U.S.N.										

B.M.S. College of Engineering, Bengaluru-560019

Autonomous Institute Affiliated to VTU

July 2023 Semester End Main Examinations

Program: B.E. **Semester: VI**
Branch: Biotechnology **Duration: 3 hrs.**
Course Code: 19BT6DCBIN **Max Marks: 100**
Course: Bioinformatics **Date: 14.07.2023**

Instructions: 1. Answer any FIVE full questions, choosing one full question from each unit.
2. Missing data, if any, may be suitably assumed.

Important Note: Completing your answers, compulsorily draw diagonal cross lines on the remaining blank pages. Revealing of identification, appeal to evaluator will be treated as malpractice.

		UNIT – I	*CO*	*PO*	Marks
1	a)	Outline the importance of KEGG databases.	*CO1*	*PO*	**06**
	b)	With representative data show the column information available at 'ATOM' row in PDB flat file.	*CO1*	*PO*	**08**
	c)	How is Uniprot-SWISSPROT unique from Uniprot-TrEMBL.	*CO1*	*PO*	**06**
		UNIT – II			
2	a)	Construct the global sequence alignment using Needleman–Wunsch algorithm for the given two sequences. Assume Match =+1, miss match=-1, gap penalty =-1 Sequence 1: GTACA Sequence 2: GTCACCA	*CO2*	*PO5*	**10**
	b)	With a suitable illustration for dot matrix method, show the alignment for perfect match, repeats, palindrome, minisatellites and indels through graphical representation.	*CO3*	*PO2*	**05**
	c)	What is the difference between PAM and BLOSUM matrices?	*CO1*	*PO*	**05**
		OR			
3	a)	Given below are two randomly generated sequences: ATGC TGC Create a score matrix & deduce the alignment of these two sequences using Needleman–Wunsch algorithm. Assume: Match score +1, Mismatch score -1, Gap penalty -2.	*CO2*	*PO5*	**10**
	b)	Construct the PSSM from a multiple sequence alignment provided. A T G T C G A A G A C T T A C T C A C G G A G G A A C C T G	*CO2*	*PO5*	**10**

<table>
<tr><td></td><td></td><td></td><td>UNIT-III</td><td></td><td></td><td></td></tr>
<tr><td>4</td><td>a)</td><td>Following is a given distance matrix comprising six taxa: Using UPGMA algorithm calculate stepwise distance matrix and construct a tree comprising six taxa.
<table><tr><td></td><td>A</td><td>B</td><td>C</td></tr><tr><td>B</td><td>0.40</td><td></td><td></td></tr><tr><td>C</td><td>0.35</td><td>0.45</td><td></td></tr><tr><td>D</td><td>0.60</td><td>0.70</td><td>0.55</td></tr></table></td><td>CO2</td><td>PO5</td><td>10</td></tr>
<tr><td></td><td>b)</td><td>Illustrate the significance of bootstrapping in phylogenetic tree construction.</td><td>CO1</td><td>PO</td><td>10</td></tr>
<tr><td></td><td></td><td>OR</td><td></td><td></td><td></td></tr>
<tr><td>5</td><td>a)</td><td>Interpret the pattern: [AC]-x-V-x(4)-{ED}.</td><td>CO3</td><td>PO2</td><td>04</td></tr>
<tr><td></td><td>b)</td><td>For the following distances between five taxa, construct a phylogenetic tree using FM method and determine the branch lengths.
<table><tr><td></td><td>A</td><td>B</td><td>C</td><td>D</td><td>E</td></tr><tr><td>A</td><td>-</td><td>12</td><td>14</td><td>14</td><td>15</td></tr><tr><td>B</td><td>-</td><td>-</td><td>12</td><td>12</td><td>13</td></tr><tr><td>C</td><td>-</td><td>-</td><td>-</td><td>6</td><td>7</td></tr><tr><td>D</td><td>-</td><td>-</td><td>-</td><td>-</td><td>3</td></tr><tr><td>E</td><td>-</td><td>-</td><td>-</td><td>-</td><td>-</td></tr></table></td><td>CO2</td><td>PO5</td><td>12</td></tr>
<tr><td></td><td>c)</td><td>For the given multiple sequence alignment, construct the HMM by representing both emission and transition probabilities.
A C A - - - A T G
T C A A C T A T C
A C A C - - A G C
A G A - - - A T C
A C C G - - A T C</td><td>CO3</td><td>PO2</td><td>04</td></tr>
<tr><td></td><td></td><td>UNIT –I V</td><td></td><td></td><td></td></tr>
<tr><td>6</td><td>a)</td><td>List and comment on the tools used in genomic and proteomics sequence acquisition and analysis.</td><td>CO1</td><td>PO</td><td>10</td></tr>
<tr><td></td><td>b)</td><td>Discuss the principle involved in identifying the restriction enzymes using bioinformatics tools.</td><td>CO1</td><td>PO</td><td>05</td></tr>
<tr><td></td><td>c)</td><td>What are the applications of structure visualization software's used for biomolecules.</td><td>CO1</td><td>PO</td><td>05</td></tr>
<tr><td></td><td></td><td>UNIT – V</td><td></td><td></td><td></td></tr>
<tr><td>7</td><td>a)</td><td>Discuss the principle involved in the protein structure prediction using homology modeling.</td><td>CO1</td><td>PO</td><td>07</td></tr>
<tr><td></td><td>b)</td><td>How the Ramachandran plot plays an important role during protein tertiary structure prediction.</td><td>CO1</td><td>PO</td><td>07</td></tr>
<tr><td></td><td></td><td>c)</td><td>Define the role of molecular docking in drug discovery? Comment on its advantages and disadvantages.</td><td>CO3</td><td>PO3</td><td>06</td></tr>
</table>

****